
Approximation of weak adjoints by reverse automatic differentiation of BDF methods

Dörte Beigel · Mario S. Mommer ·
Leonard Wirsching · Hans Georg Bock

September 13, 2011

Abstract With this contribution, we shed light on the relation between the discrete adjoints of multistep backward differentiation formula (BDF) methods and the solution of the adjoint differential equation. To this end, we develop a functional-analytic framework based on a constrained variational problem and introduce the notion of weak adjoint solutions. We devise a finite element Petrov-Galerkin interpretation of the BDF method together with its discrete adjoint scheme obtained by reverse internal numerical differentiation. We show how the finite element approximation of the weak adjoint is computed by the discrete adjoint scheme and prove its asymptotic convergence in the space of normalized functions of bounded variation. We also obtain asymptotic convergence of the discrete adjoints to the classical adjoints on the inner time interval. Finally, we give numerical results for non-adaptive and fully adaptive BDF schemes. The presented framework opens the way to carry over the existing theory on global error estimation techniques from finite element methods to BDF methods.

Keywords BDF methods · discrete adjoints · Petrov-Galerkin discretization

Mathematics Subject Classification (2000) 65L06 · 65L60 · 49K40 · 65L20

1 Introduction

Consider a nonlinear initial value problem (IVP) in ordinary differential equations (ODE) with sufficiently smooth right hand side $\mathbf{f} : [t_s, t_f] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$

Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University. Im Neuenheimer Feld 368, 69120 Heidelberg, Germany. Fax: +49-6221-54 5444.
E-mail: doerte.beigel@iwr.uni-heidelberg.de, Tel.: +49-6221-54 8896 (corresponding).
E-mail: {mario.mommer, leonard.wirsching, bock}@iwr.uni-heidelberg.de.

$$\dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t)), \quad t \in (t_s, t_f] \quad (1a)$$

$$\mathbf{y}(t_s) = \mathbf{y}_s. \quad (1b)$$

Consider also a differentiable criterion of interest J depending on the final state $\mathbf{y}(t_f)$ of the solution of (1). This is relevant whenever one is not interested in the whole solution trajectory $\mathbf{y}(t)$ or even the final state $\mathbf{y}(t_f)$, but only in a functional output of these quantities. Note that by standard reformulations (cf. [13, p.93], [6, p.25]) this setting also captures the cases of a parameter-dependent right hand side $\mathbf{f}(t, \mathbf{y}, \mathbf{p})$ and a criterion of interest of Bolza type $J(\mathbf{y}) = \int_{t_s}^{t_f} J_1(\mathbf{y}(t), \mathbf{p})dt + J_2(\mathbf{y}(t_f))$.

The adjoint differential equation corresponding to the evaluation of $J(\mathbf{y}(t_f))$ in the solution of (1) is (see Section 2)

$$\dot{\boldsymbol{\lambda}}(t) = -\mathbf{f}_{\mathbf{y}}^{\top}(t, \mathbf{y}(t))\boldsymbol{\lambda}(t), \quad t \in (t_f, t_s] \quad (2a)$$

$$\boldsymbol{\lambda}(t_f) = J'(\mathbf{y}(t_f))^{\top}. \quad (2b)$$

The adjoint solution describes the dependency of $J(\mathbf{y}(t_f))$ on disturbances of the nominal solution $\mathbf{y}(t)$. Therefore, it is of great importance in the solution of optimal control problems. For example, in indirect approaches based on the Pontryagin minimum principle, (2) appears as part of the optimality conditions.

For an approximation of the solution of (1), the solution of the adjoint differential equation (2) can be computed in two different ways, the continuous adjoint approach or the discrete adjoint approach. The former solves the adjoint differential equation by numerical integration, see for example [11]. Whereas the latter applies automatic differentiation techniques to the numerical integration scheme. This approach, firstly presented in [7], is known as internal numerical differentiation (IND). It has significant advantages in direct derivative-based approaches for the solution of optimal control problems that use integrators, e.g. direct single and multiple shooting.

In the case of Runge-Kutta methods, the discrete adjoint scheme generated by adjoint IND is itself a Runge-Kutta scheme for the adjoint differential equation (2), and thus gives a convergent approximation to the adjoint solution [7, 24]. In the case of continuous and discontinuous Galerkin methods applied to (1), the discrete adjoint scheme yields an approximation to the solution of (2) (see e.g. [18]). The discrete adjoints of discontinuous Galerkin methods for compressible Navier-Stokes equations are, for example, considered in [14].

The situation becomes significantly more complex in the case of multistep methods, as the discrete adjoint schemes of linear multistep methods (LMM) are generally *not* consistent with the adjoint differential equation (2). But they still provide approximations of the sensitivities $J'(\mathbf{y}(t_f)) \frac{\partial \mathbf{y}(t_f)}{\partial \mathbf{y}_s}$ at the initial time t_s that converge with the rate of the nominal LMM [8, 22]. Due to this property, the multistep BDF method and its discrete adjoint scheme are used successfully in direct methods for the solution of optimal control problems, e.g. in direct multiple shooting [9, 2].

In this contribution, we focus on the relation between the discrete adjoints of variable-order variable-stepsizes BDF methods and the adjoints defined by (2). To this end, we construct a suitable constrained variational problem (CVP) in a Banach space setting using the duality pairing between the space of continuous functions and its dual, the space of normalized functions of bounded variation. It turns out that the adjoint of a stationary point of this CVP is the normalized integral of the solution of the Hilbert space adjoint differential equation (2). Motivated by PDE nomenclature, we will call it a weak solution of (2) or shortly weak adjoint. We apply Petrov-Galerkin techniques, and show that with the appropriate choice of basis functions the infinite-dimensional optimality conditions of the CVP are approximated by the BDF method and its discrete adjoint scheme obtained by adjoint internal numerical differentiation of the nominal BDF scheme. In particular, we obtain that discretization and optimization commute in this Banach space setting. Finally, we prove that the finite element approximation of the weak adjoint, which can be computed by a simple post-processing of the discrete adjoints, converges to the weak adjoint on the entire time interval. This result is based on the linear convergence of the discrete adjoints to the solution of (2) on the inner time interval which is shown as well.

This paper is organized as follows. In Section 2 we derive the adjoint differential equation as part of the optimality conditions of an infinite-dimensional constrained variational problem in Hilbert spaces. The BDF method and its discrete adjoint scheme generated by internal numerical differentiation techniques are then described in Section 3. In Section 4 we present the optimality conditions of the constrained variational problem embedded into the Banach space of all continuously differentiable functions. After showing the well-posedness of the optimality conditions and their relation to the Hilbert space optimality conditions, we extend the setting to capture the space of all functions that are continuous and piecewise continuously differentiable. For the Petrov-Galerkin discretization of Section 5 we choose suitable finite element spaces that yield equivalence between the discretized optimality conditions and the BDF scheme together with its discrete adjoint scheme. In Section 6 we start by proving the convergence of the discrete adjoints to the solution of the Hilbert space adjoint equation on the inner time interval. Using this result, we show the convergence of the finite element approximation to the weak adjoint solution. Section 7 presents numerical results on a nonlinear test case with analytic solutions.

2 Initial value problems and their adjoints in a Hilbert space setting

In this section, we derive the adjoint differential equation in a Hilbert space functional-analytic setting. Our goal is to specify the assumptions on the initial value problem, to settle some notation, and to lay the groundwork for the constructions that follow. In particular, we make explicit the connection

between the adjoint differential equation and the Lagrange multiplier of the solution of a constrained variational problem in a Hilbert space setting based on the Sobolev spaces usually found in finite element formulations.

2.1 Existence, uniqueness and differentiability of the nominal solution

Assume that the right hand side $\mathbf{f}(t, \mathbf{y})$ of (1) is continuous on an open set $\mathcal{D} \subset \mathbb{R} \times \mathbb{R}^d$ with $(t_s, \mathbf{y}_s) \in \mathcal{D}$ and its first-order partial derivative $\mathbf{f}_{\mathbf{y}}(t, \mathbf{y})$ is continuous on \mathcal{D} . Thus, according to the Picard-Lindelöf Theorem [13], problem (1) is well-posed in the sense of Hadamard, i.e. it admits a unique solution depending continuously on the input data. Beyond that, the solution $\mathbf{y}(t)$ is continuously differentiable on an open interval \mathcal{I} , see [13], and we assume that t_f is chosen such that $[t_s, t_f] \subset \mathcal{I}$. Thus, the solution $\mathbf{y}(t)$ of (1) lies in the Banach space $C^1[t_s, t_f]^d$ of all continuously differentiable functions from $[t_s, t_f]$ to \mathbb{R}^d equipped with the usual norm. Furthermore, the solution $\mathbf{y}(t) = \mathbf{y}(t; t_s, \mathbf{y}_s)$ is continuously differentiable with respect to \mathbf{y}_s and the derivatives $\mathbf{w}_i(t) = \partial \mathbf{y}(t; t_s, \mathbf{y}_s) / \partial (\mathbf{y}_s)_i$ solve [13]

$$\dot{\mathbf{w}}_i(t) = \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t)) \mathbf{w}_i(t), \quad t \in (t_s, t_f] \quad (3a)$$

$$\mathbf{w}_i(t_s) = \mathbf{e}_i \quad (3b)$$

where \mathbf{e}_i is the i th unit vector, $i \in \{1, \dots, d\}$. Moreover, $\mathbf{w}_i(t)$ exists uniquely and is continuously differentiable on $[t_s, t_f]$ since the partial derivative of the right hand side of (3) with respect to \mathbf{w}_i is continuous in (t, \mathbf{w}_i) . The residual of (1a)

$$\rho(\mathbf{y}) := \dot{\mathbf{y}}(\cdot) - \mathbf{f}(\cdot, \mathbf{y}(\cdot)) \quad (4)$$

lies in the Banach space $C^0[t_s, t_f]^d$ of all continuous functions from $[t_s, t_f]$ to \mathbb{R}^d equipped with the standard norm $\|\mathbf{g}\|_{C^0[t_s, t_f]^d} = \sum_{i=1}^d \|g_i\|_{C^0[t_s, t_f]}$ where $\|g_i\|_{C^0[t_s, t_f]} = \max_{t \in [t_s, t_f]} |g_i(t)|$.

2.2 Lagrange multipliers and adjoint differential equations

The core of this section is the identification of the adjoint as the Lagrange multiplier of a constrained optimization problem in a functional-analytic setting. The ideas described here are of course not new. However, the setting for the case of ordinary differential equations is fundamental for this contribution. Since we have not found it in the literature, we include here a detailed derivation.

Recall that functions in $C^0[t_s, t_f]^d$, restricted to the open interval (t_s, t_f) , form a dense subset of the space $L^2(t_s, t_f)^d$ of all quadratically Lebesgue-integrable functions. Similarly, recall that the subset $C^1[t_s, t_f]^d$ is dense in the Sobolev space $H^1(t_s, t_f)^d$ of all $L^2(t_s, t_f)^d$ -functions with weak derivative in

$L^2(t_s, t_f)^d$ (see [1, Ch.3]). Furthermore, both spaces $L^2(t_s, t_f)^d$ and $H^1(t_s, t_f)^d$ are Hilbert spaces.

Knowing this, we embed the initial value problem (1) into an optimization framework and derive the adjoint differential equation as part of the first-order necessary optimality conditions. To this end, we consider the constrained variational problem

$$\min_{\mathbf{y}} J(\mathbf{y}(t_f)) \quad (5a)$$

$$\text{s. t. } \dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t)), \quad t \in (t_s, t_f] \quad (5b)$$

$$\mathbf{y}(t_s) = \mathbf{y}_s \quad (5c)$$

which is equivalent to evaluating $J(\mathbf{y}(t_f))$ in the solution of (1). Considering (5) on the space $H^1(t_s, t_f)^d$, the Hilbert space Lagrangian $\mathcal{L} : H^1(t_s, t_f)^d \times L^2(t_s, t_f)^d \rightarrow \mathbb{R}$ of (5) using the L^2 -scalar product is

$$\mathcal{L}(\mathbf{y}, \boldsymbol{\lambda}) := J(\mathbf{y}(t_f)) - \int_{t_s}^{t_f} \boldsymbol{\lambda}^\top(t) [\dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t))] dt - \boldsymbol{\lambda}^\top(t_s) [\mathbf{y}(t_s) - \mathbf{y}_s]$$

where $\boldsymbol{\lambda}$ is the Lagrange multiplier. The optimality condition of (5) is based on the Fréchet derivative of \mathcal{L} at $(\mathbf{y}, \boldsymbol{\lambda})$ in direction $(\mathbf{w}, \boldsymbol{\chi})$ which exists due to Fréchet differentiability of J and [16, Ch.0§0.2.5]

$$\begin{aligned} \mathcal{L}'(\mathbf{y}, \boldsymbol{\lambda})(\mathbf{w}, \boldsymbol{\chi}) &= \mathcal{L}_{\mathbf{y}}(\mathbf{y}, \boldsymbol{\lambda})(\mathbf{w}) + \mathcal{L}_{\boldsymbol{\lambda}}(\mathbf{y}, \boldsymbol{\lambda})(\boldsymbol{\chi}) \\ &= \left\{ J'(\mathbf{y}(t_f))\mathbf{w}(t_f) - \int_{t_s}^{t_f} \boldsymbol{\lambda}^\top(t) [\dot{\mathbf{w}}(t) - \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t))\mathbf{w}(t)] dt - \boldsymbol{\lambda}^\top(t_s)\mathbf{w}(t_s) \right\} \\ &\quad + \left\{ - \int_{t_s}^{t_f} \boldsymbol{\chi}^\top(t) [\dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t))] dt - \boldsymbol{\chi}^\top(t_s) [\mathbf{y}(t_s) - \mathbf{y}_s] \right\}. \end{aligned}$$

The necessary condition for a stationary point $(\mathbf{y}, \boldsymbol{\lambda}) \in H^1(t_s, t_f)^d \times L^2(t_s, t_f)^d$ of (5) is that $\mathcal{L}'(\mathbf{y}, \boldsymbol{\lambda})(\mathbf{w}, \boldsymbol{\chi}) = 0$ holds for all directions $(\mathbf{w}, \boldsymbol{\chi}) \in H^1(t_s, t_f)^d \times L^2(t_s, t_f)^d$. Choosing $\mathbf{w} = \mathbf{0} \in H^1(t_s, t_f)^d$ and only varying $\boldsymbol{\chi} \in L^2(t_s, t_f)^d$ the necessary condition reads

$$\int_{t_s}^{t_f} \boldsymbol{\chi}^\top(t) [\dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t))] dt + \boldsymbol{\chi}^\top(t_s) [\mathbf{y}(t_s) - \mathbf{y}_s] = 0, \quad \forall \boldsymbol{\chi} \quad (6)$$

which possesses the same unique solution $\mathbf{y} \in C^1[t_s, t_f]^d$ as (1). Taking now $\boldsymbol{\chi} = \mathbf{0} \in L^2(t_s, t_f)^d$ and only varying $\mathbf{w} \in H^1(t_s, t_f)^d$ one obtains using integration by parts

$$[J'(\mathbf{y}(t_f)) - \boldsymbol{\lambda}^\top(t_f)] \mathbf{w}(t_f) - \int_{t_f}^{t_s} \left[\dot{\boldsymbol{\lambda}}(t) + \mathbf{f}_{\mathbf{y}}^\top(t, \mathbf{y}(t))\boldsymbol{\lambda}(t) \right]^\top \mathbf{w}(t) dt = 0, \quad \forall \mathbf{w}$$

which possesses the same solution as (2). Under the assumptions of Section 2.1, the unique solution $\boldsymbol{\lambda}(t)$ of (2) is continuously differentiable on $[t_s, t_f]$ and depends continuously on $J'(\mathbf{y}(t_f))^\top$.

3 Efficient solution of initial value problems and sensitivity generation

We now review the numerical solution of ODEs using BDF methods, and the corresponding sensitivity generation using automatic differentiation techniques. We briefly introduce BDF methods with an emphasis on the trajectories they define as functions of time. Then, we show how to obtain discrete adjoints in the BDF context, and review what is known so far about their relation to the solution of (2).

3.1 Backward differentiation formula method

This section follows the lines of [23, p.181ff and p.253f]. Consider the backward differentiation formula method

$$\mathbf{y}_0 = \mathbf{y}_s \quad (7a)$$

$$\sum_{i=0}^{k_n} \alpha_i^{(n)} \mathbf{y}_{n+1-i} = h_n \mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}), \quad n = 0, \dots, N-1 \quad (7b)$$

with a self-starting procedure that begins with $k_0 = 1$ (implicit Euler) and increases successively the order of the steps until the maximum order is reached. Note that BDF methods are used up to order 6, since for higher order they become unstable. In practical implementations both the stepsize h_n and the order k_n are chosen adaptively to obtain better performance. The numerical solution is computed at discrete time points $t_s = t_0 < t_1 < \dots < t_N = t_f$ with $t_{n+1} = t_n + h_n$ and \mathbf{y}_n denotes the numerical approximation to the value $\mathbf{y}(t_n)$. The coefficients $\alpha_i^{(n)}$ are determined by

$$\alpha_i^{(n)} = h_n \dot{L}_i^{(n)}(t_{n+1}), \text{ where } L_i^{(n)}(t) = \prod_{j=0, j \neq i}^{k_n} \frac{t - t_{n+1-j}}{t_{n+1-i} - t_{n+1-j}} \quad (8)$$

are the fundamental Lagrangian polynomials. Thus, the coefficients depend on the discrete time points and the order. In each step, the BDF method provides a polynomial approximation to the solution $\mathbf{y}(t)$ of (1) in a natural way through the interpolation polynomial

$$\mathbf{y}(t)|_{t \in [t_n, t_{n+1}]} \approx \sum_{i=0}^{k_n} L_i^{(n)}(t) \mathbf{y}_{n+1-i}, \quad (9)$$

also known as *dense output*. The composition of all these polynomials gives a continuous and piecewise continuously differentiable approximation to the solution $\mathbf{y}(t)$ on the whole time interval $[t_s, t_f]$.

3.2 Adjoint differentiation of BDF integration schemes

The basic idea of internal numerical differentiation [7] is to differentiate the discretization scheme used to obtain the nominal approximations $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_N$ for specified adaptive components h_n and k_n using automatic differentiation (AD) techniques either in forward or in adjoint mode. Adjoint IND was first described in [8] for Runge-Kutta integration schemes and later on in [10] for BDF methods. Applying adjoint IND to the BDF scheme (7) we obtain the discrete adjoint scheme

$$\alpha_0^{(N-1)} \boldsymbol{\lambda}_N - J'(\mathbf{y}_N)^\top = h_{N-1} \mathbf{f}_\mathbf{y}^\top(t_N, \mathbf{y}_N) \boldsymbol{\lambda}_N \quad (10a)$$

$$\sum_{\substack{0 \leq i \leq N-1-n \\ i \leq k_{\max}}} \alpha_i^{(n+i)} \boldsymbol{\lambda}_{n+1+i} = h_n \mathbf{f}_\mathbf{y}^\top(t_{n+1}, \mathbf{y}_{n+1}) \boldsymbol{\lambda}_{n+1}, \quad n = N-2, \dots, 0 \quad (10b)$$

with input direction $J'(\mathbf{y}_N)^\top$ and the convention $\alpha_i^{(n)} = 0$ for $i > k_n$, $k_{\max} = \max_n \{k_n\}$ (see also [22]). This scheme forms together with (7) the optimality conditions of the nonlinear program (NLP)

$$\min_{\mathbf{y}} J(\mathbf{y}_N) \quad \text{s. t.} \quad (7) \quad (11)$$

with $\mathbf{y}^\top := [\mathbf{y}_0^\top \mathbf{y}_1^\top \dots \mathbf{y}_N^\top]$. This NLP is a discretization of the constrained variational problem (5).

The discrete adjoints given by (10) are the exact derivatives of the nominal integration scheme (7) (beside round-off errors). Furthermore, for a BDF scheme with constant order k , the discrete adjoint $\boldsymbol{\lambda}_1$ converges with the same order k to the value $\boldsymbol{\lambda}(t_s)$ of the adjoint solution of (2), cf. [8, 22].

The discrete adjoints are generally inconsistent approximations to the solution of (2) around a nominal approximation passing through $\{\mathbf{y}_n\}_{n=0}^N$, see Figure 1(b). In the case of constant order k and constant stepsizes h , the discrete adjoints coming from the adjoint initialization and adjoint termination are inconsistent as well, whereas the main part, i.e. formula (10b) with $n = N-k, \dots, k$, gives consistent approximations of order k , see Figure 1(a).

Due to the inconsistency of the discrete adjoint scheme (10) with the adjoint differential equation (2) discretization and optimization of (5) do not commute in the commonly used Hilbert space setting. This gives rise to the question for a new functional-analytic setting that is suitable for multistep methods. The next sections are devoted to the development of this setting.

4 Solution of the constrained variational problem in a Banach space setting

As seen in the previous section, the Hilbert space setting of Section 2 is not suitable to analyze multistep methods and their discrete adjoints. Here, we propose to embed the constrained variational problem (5) into a Banach space setting and show the well-posedness of the corresponding infinite-dimensional optimality conditions.

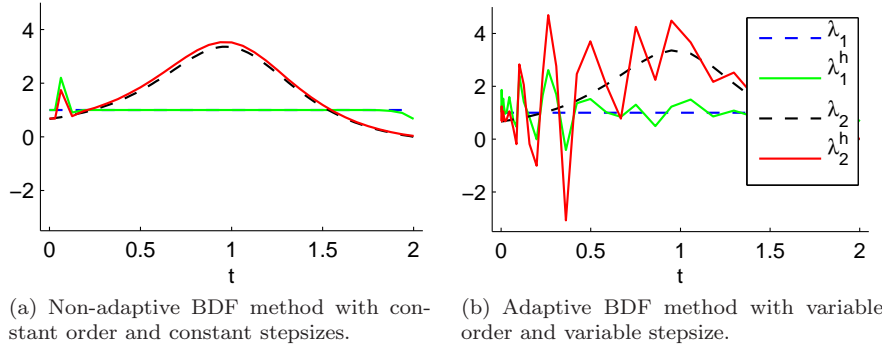


Fig. 1 Comparison of discrete adjoints $\lambda^h = [\lambda_1^h, \lambda_2^h]^\top$ and analytic solution $\lambda = [\lambda_1, \lambda_2]^\top$ of the Hilbert space adjoint differential equation on the Catenary test case (see Section 7).

4.1 General considerations

Duality pairing According to Section 2.1, the residual $\rho(\mathbf{y})$ of (1a) is an element of the space $C^0[t_s, t_f]^d$. Thus, we focus on the duality pairing between the Banach space $C^0[t_s, t_f]^d$ and its dual. The Riesz Representation Theorem [20, Ch.5§5.5] states that for every continuous linear functional \mathfrak{L} on $C^0[t_s, t_f]$ exists a unique $\Psi \in \text{NBV}[t_s, t_f]$ such that

$$\mathfrak{L}[g] = \langle \Psi, g \rangle_{\text{NBV}[t_s, t_f], C^0[t_s, t_f]} = \int_{t_s}^{t_f} g(t) d\Psi(t), \quad (12)$$

where the integral is the Riemann-Stieltjes integral [21, Ch.VIII§6]. The Banach space $\text{NBV}[t_s, t_f]$ consists of all normalized functions of bounded variation on $[t_s, t_f]$ that are zero in t_s and continuous from the right on (t_s, t_f) . It is equipped with the total variation norm

$$\|\Psi\|_{\text{NBV}[t_s, t_f]} = \sup \sum_{i=1}^m |\Psi(t_i) - \Psi(t_{i-1})|$$

where the supremum is taken over all partitions $t_s = t_0 < \dots < t_m = t_f$ of $[t_s, t_f]$. According to the Riesz Representation Theorem, for each Ψ the value of the total variation norm coincides with the value of the dual norm given by

$$\|\Psi\|_{\text{NBV}[t_s, t_f]} = \max_{\|g\|_{C^0[t_s, t_f]}=1} |\langle \Psi, g \rangle_{\text{NBV}[t_s, t_f], C^0[t_s, t_f]}|.$$

Hence, we will always use the norm that is better suited in the particular situation. The dual of the finite Cartesian product $C^0[t_s, t_f]^d$ is the finite Cartesian product $\text{NBV}[t_s, t_f]^d$ of the duals with duality pairing

$$\langle \Psi, \mathbf{g} \rangle_{\text{NBV}^d, (C^0)^d} = \sum_{i=1}^d \langle \Psi_i, g_i \rangle_{\text{NBV}, C^0} = \sum_{i=1}^d \int_{t_s}^{t_f} g_i(t) d\Psi_i(t) =: \int_{t_s}^{t_f} \mathbf{g}(t) d\Psi(t)$$

and dual norm $\|\Psi\|_{\text{NBV}[t_s, t_f]^d} = \max_{1 \leq i \leq d} \|\Psi_i\|_{\text{NBV}[t_s, t_f]}$, see [26, Ch.II§12.1].

Variational formulation of the initial value problem The variational formulation of (1) on the described Banach spaces reads: Find $\mathbf{y} \in C^1[t_s, t_f]^d$ with $\mathbf{y}(t_s) = \mathbf{y}_s$ such that

$$\int_{t_s}^{t_f} \dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t)) \, d\mathbf{\Gamma}(t) = 0 \quad \forall \mathbf{\Gamma} \in \text{NBV}[t_s, t_f]^d. \quad (13)$$

This problem possesses at least one solution which is the strong solution given by (1). The uniqueness follows from the fact that for continuous functions $g \in C^0[t_s, t_f]$ it holds

$$\int_{t_s}^{t_f} g(t) \, d\Psi(t) = 0 \quad \forall \Psi \in \text{NBV}[t_s, t_f] \quad \Rightarrow \quad g = 0.$$

Thus, both formulations (1) and (13) give the same solution $\mathbf{y}(t)$ and (13) is well-posed according to the well-posedness of (1).

4.2 Infinite-dimensional optimality conditions

Considering the constrained variational problem (5) on the function space $C^1[t_s, t_f]^d$, the Lagrangian $\mathcal{L} : C^1[t_s, t_f]^d \times \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is given by

$$\mathcal{L}(\mathbf{y}, \mathbf{\Lambda}, \mathbf{l}) := J(\mathbf{y}(t_f)) - \int_{t_s}^{t_f} \dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t)) \, d\mathbf{\Lambda}(t) - \mathbf{l}^\top [\mathbf{y}(t_s) - \mathbf{y}_s] \quad (14)$$

where the Lagrange multipliers \mathbf{l} and $\mathbf{\Lambda}$ lie in the corresponding dual spaces \mathbb{R}^d and $\text{NBV}[t_s, t_f]^d$. The Lagrangian is based on the variational formulation (13) and includes the initial condition using an additional Lagrange multiplier. We first state the central theorem of this section which describes the stationary point of \mathcal{L} and defer the proof for the end of the section.

Theorem 1 *The optimality conditions of the constrained variational problem (5) on $C^1[t_s, t_f]^d$, i.e.*

$$J'(\mathbf{y}(t_f))\mathbf{w}(t_f) - \int_{t_s}^{t_f} \dot{\mathbf{w}}(t) - \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t))\mathbf{w}(t) \, d\mathbf{\Lambda}(t) - \mathbf{l}^\top \mathbf{w}(t_s) = 0, \quad (15a)$$

$$- \int_{t_s}^{t_f} \dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t)) \, d\mathbf{\Gamma}(t) = 0, \quad (15b)$$

$$-\mathbf{r}^\top [\mathbf{y}(t_s) - \mathbf{y}_s] = 0, \quad (15c)$$

$$\forall (\mathbf{w}, \mathbf{\Gamma}, \mathbf{r}) \in C^1[t_s, t_f]^d \times \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d,$$

possess a unique solution $(\mathbf{y}, \mathbf{\Lambda}, \mathbf{l})$ in $C^1[t_s, t_f]^d \times \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d$. Moreover, $\mathbf{y}(t)$ is the solution of (1), and \mathbf{l} and $\mathbf{\Lambda}(t)$ are given in terms of the adjoint solution $\boldsymbol{\lambda}(t)$ of (2)

$$\mathbf{l} = \boldsymbol{\lambda}(t_s), \quad \mathbf{\Lambda}(t) = \int_{t_s}^t \boldsymbol{\lambda}(\tau) \, d\tau, \quad (16)$$

with componentwise integration.

The necessary optimality condition for a stationary point $(\mathbf{y}, \mathbf{A}, \mathbf{l})$ of the Lagrangian (14) is given by

$$\begin{pmatrix} \mathcal{L}_{\mathbf{y}}(\mathbf{y}, \mathbf{A}, \mathbf{l})(\mathbf{w}) \\ \mathcal{L}_{\mathbf{A}}(\mathbf{y}, \mathbf{A}, \mathbf{l})(\mathbf{r}) \\ \mathcal{L}_{\mathbf{l}}(\mathbf{y}, \mathbf{A}, \mathbf{l})(\mathbf{r}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \forall \mathbf{w} \in C^1[t_s, t_f]^d, \quad \mathbf{r} \in \mathbb{R}^d$$

which is exactly (15). As equations (15b)-(15c) are already given by (13) and discussed over there, we now focus on equation (15a) of the optimality conditions. Provided that $\mathbf{y}(t)$ is known, the adjoint problem in variational formulation reads: Find $(\mathbf{A}, \mathbf{l}) \in \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d$ such that (15a) holds for all $\mathbf{w} \in C^1[t_s, t_f]^d$.

Lemma 1 *For the solution $\mathbf{y}(t)$ of (15b)-(15c), a corresponding adjoint solution $(\mathbf{A}, \mathbf{l}) \in \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d$ of (15a) is provided by (16).*

Proof Recall that the adjoint differential equation (2) has a unique solution $\boldsymbol{\lambda} \in C^1[t_s, t_f]^d$ (cf. Section 2.2). Multiplying the transposed of (2a) from the right by any $\mathbf{w} \in C^1[t_s, t_f]^d$, integrating over $[t_s, t_f]$ and adding the transposed of (2b) multiplied by $\mathbf{w}(t_f)$ yields

$$\int_{t_s}^{t_f} [\dot{\boldsymbol{\lambda}}(t) + \mathbf{f}_{\mathbf{y}}^T(t, \mathbf{y}(t))\boldsymbol{\lambda}(t)]^T \mathbf{w}(t) dt - [\boldsymbol{\lambda}(t_f) - J'(\mathbf{y}(t_f))]^T \mathbf{w}(t_f) = 0. \quad (17)$$

Integration by parts gives for all $\mathbf{w} \in C^1[t_s, t_f]^d$

$$\int_{t_s}^{t_f} \boldsymbol{\lambda}^T(t) [\dot{\mathbf{w}}(t) - \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t))\mathbf{w}(t)] dt - \boldsymbol{\lambda}^T(t_s)\mathbf{w}(t_s) + J'(\mathbf{y}(t_f))\mathbf{w}(t_f) = 0.$$

Consequently, (16) provides a solution $(\mathbf{A}, \mathbf{l}) \in \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d$ of (15a), since the indefinite integral $\Lambda_i(t) = \int_{t_s}^t \lambda_i(\tau) d\tau$ is a normalized function of bounded variation [19, Sec.32] and it holds $\int_{t_s}^{t_f} g(t) d\Lambda_i(t) = \int_{t_s}^{t_f} \Lambda_i'(t) g(t) dt = \int_{t_s}^{t_f} \lambda_i(t) g(t) dt$, cf. [21, Ch.VIII§6]. \square

The next lemma proves the uniqueness of the weak adjoint solution.

Lemma 2 *For the solution $\mathbf{y}(t)$ of (15b)-(15c), the corresponding adjoint solution $(\mathbf{A}, \mathbf{l}) \in \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d$ of (15a) is unique.*

Proof Equation (15a) is equivalent to

$$\underbrace{\int_{t_s}^{t_f} \dot{\mathbf{w}}(t) - \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t))\mathbf{w}(t) d\mathbf{A}(t) + \mathbf{l}^T \mathbf{w}(t_s)}_{=: \mathbf{A}(\mathbf{A}, \mathbf{l})(\mathbf{w})} = \underbrace{J'(\mathbf{y}(t_f))\mathbf{w}(t_f)}_{=: B(\mathbf{w})} \quad \forall \mathbf{w} \in C^1[t_s, t_f]^d$$

where B and $\mathbf{A}(\mathbf{A}, \mathbf{l})$ are linear functionals on $C^1[t_s, t_f]^d$ and $\mathbf{A} : \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d \rightarrow (C^1[t_s, t_f]^d)'$ is linear in (\mathbf{A}, \mathbf{l}) . We have to show that $\mathcal{N}(\mathbf{A}) = \{(\mathbf{0}, \mathbf{0})\}$, where the nullspace of \mathbf{A} is given by

$$\mathcal{N}(\mathbf{A}) = \{(\mathbf{A}, \mathbf{l}) \in \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d : \mathbf{A}(\mathbf{A}, \mathbf{l})(\mathbf{w}) = 0 \quad \forall \mathbf{w} \in C^1[t_s, t_f]^d\}.$$

Due to Section 2.1, for every initial value $\mathbf{w}_1(t_s) \in \mathbb{R}^d$ there exists a function $\mathbf{w}_1 \in C^1[t_s, t_f]^d$ that satisfies (3a). Inserting \mathbf{w}_1 in $\mathbf{A}(\mathbf{A}, \mathbf{l})$ then gives

$$\mathbf{A}(\mathbf{A}, \mathbf{l})(\mathbf{w}_1) = \int_{t_s}^{t_f} \mathbf{0} \, d\mathbf{A}(t) + \mathbf{l}^\top \mathbf{w}_1(t_s) = 0 + \mathbf{l}^\top \mathbf{w}_1(t_s).$$

Thus, \mathbf{l} has to vanish in order to ensure $\mathbf{A}(\mathbf{A}, \mathbf{l})(\mathbf{w}) = 0 \, \forall \mathbf{w} \in C^1[t_s, t_f]^d$. Now, we search for functions $\mathbf{A} \in \text{NBV}[t_s, t_f]^d$ with

$$\mathbf{A}(\mathbf{A}, \mathbf{0})(\mathbf{w}) = \int_{t_s}^{t_f} \dot{\mathbf{w}}(t) - \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t))\mathbf{w}(t) \, d\mathbf{A}(t) = 0 \quad \forall \mathbf{w} \in C^1[t_s, t_f]^d.$$

With $\mathbf{g}(t) := \dot{\mathbf{w}}(t) - \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t))\mathbf{w}(t)$, it is the same to vary either $\mathbf{w} \in C^1[t_s, t_f]^d$ or $\mathbf{g} \in C^0[t_s, t_f]^d$, since the inhomogeneous ODE possesses a unique solution $\mathbf{w}(t)$ for every $\mathbf{g}(t)$. According to the uniqueness of Ψ in (12) it holds

$$\int_{t_s}^{t_f} \mathbf{g}(t) \, d\mathbf{A}(t) = 0 \quad \forall \mathbf{g} \in C^0[t_s, t_f]^d \quad \Rightarrow \quad \mathbf{A} = \mathbf{0}.$$

Consequently, $\mathcal{N}(\mathbf{A}) = \{(\mathbf{0}, \mathbf{0})\}$ which proves the uniqueness of the solution of (15a). \square

With this knowledge at hand we can now come to the proof of Theorem 1.

Proof (of Theorem 1) As seen in Section 4.1, the equations (15b)-(15c) have the same unique solution $\mathbf{y}(t)$ as (1) which implies their well-posedness. According to Lemma 1, a solution of (15a) is provided by (16). Furthermore, it is the only solution of (15a) according to Lemma 2. Since $\boldsymbol{\lambda}(t)$ depends continuously on $J'(\mathbf{y}(t_f))^\top$ (cf. Section 2.2) this still holds for $\mathbf{A}(t)$ and \mathbf{l} . Thus, (15a) together with (15b)-(15c) is well-posed. \square

With the concept of weak solutions from partial differential equations (see e.g. [17]), the triple $(\mathbf{y}, \mathbf{A}, \mathbf{l})$ is a weak solution of (1) and (2), since it solves the variational formulation (15) of (1) and (2). Thus, we will call \mathbf{A} a *weak adjoint solution* of (2) or shortly *weak adjoint*. Note that for the nominal solution, the weak solution \mathbf{y} defined by (15c)-(15b) is directly the classical solution of (1). Whereas for the adjoint, the weak solution \mathbf{A} is sufficiently regular such that a classical solution of (2) is provided by $\mathbf{A}' = \boldsymbol{\lambda}$.

4.3 Extension of the infinite-dimensional optimality conditions

As seen in Section 3.1 the approximations to the solution of (1) obtained from BDF methods are not continuously differentiable on the whole interval $[t_s, t_f]$ but rather continuous and piecewise continuously differentiable. To capture this case, an appropriate extension of the trial space $C^1[t_s, t_f]^d$ is required. To this end, we employ a time grid $t_s = t_0 < t_1 < \dots < t_N = t_f$ and a partition

of $[t_s, t_f]$ using subintervals $I_n = (t_n, t_{n+1}]$ of length $h_n = t_{n+1} - t_n$ such that $[t_s, t_f] = \{t_s\} \cup I_0 \cup \dots \cup I_{N-1}$. Choosing the trial space as

$$Y[t_s, t_f]^d := \left\{ \mathbf{y} \in C^0[t_s, t_f]^d : \mathbf{y}|_{I_n} \in C_b^1(I_n)^d \right\}, \quad (18)$$

where $C_b^1(I_n)$ is the space of all continuously differentiable and bounded functions with bounded derivative [1, Ch.1], the extended Lagrangian $\hat{\mathcal{L}} : Y[t_s, t_f]^d \times \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ of (5) solved on the function space $Y[t_s, t_f]^d$ is

$$\hat{\mathcal{L}}(\mathbf{y}, \mathbf{A}, \mathbf{l}) := J(\mathbf{y}(t_f)) - \sum_{n=0}^{N-1} \int_{I_n} \dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t)) \, d\mathbf{A}(t) - \mathbf{l}^\top [\mathbf{y}(t_s) - \mathbf{y}_s].$$

The Lagrangian $\hat{\mathcal{L}}$ is based on the extension $\hat{\mathfrak{L}}$ of the linear functional \mathfrak{L} given by (12) from $C^0[t_s, t_f]$ to $Y[t_s, t_f]$. The existence of $\hat{\mathfrak{L}}$ is guaranteed due to [26, p. 89]. We define the extended Riemann-Stieltjes integral on $I_n = (t_n, t_{n+1}]$ using the partition $t_n < \tau_0 < \tau_1 < \dots < \tau_m = t_{n+1}$ and the convention that $\theta_k = \tau_{k-1} \in [\tau_{k-1}, \tau_k]$ for $k = 1, \dots, m$ by

$$\int_{(t_n, t_{n+1}]} g(t) d\Psi(t) = \lim_{m \rightarrow \infty} \sum_{k=1}^m g(\tau_k) [\Psi(\tau_k) - \Psi(\tau_{k-1})] \quad (19)$$

such that

$$\hat{\mathfrak{L}}[g] = \sum_{n=0}^{N-1} \int_{I_n} g(t) d\Psi(t).$$

This extension $\hat{\mathfrak{L}}$ restricted to the continuous functions $g \in C^0[t_s, t_f]$ coincides with \mathfrak{L} . Thus, the same holds for the Lagrangian $\hat{\mathcal{L}}$. Furthermore, if $g \in C^0[t_n, t_{n+1}]$ then $\int_{t_n}^{t_{n+1}} g(t) d\Psi(t) = \int_{I_n} g(t) d\Psi(t)$.

With these definitions at hand, we first state the main result of the section.

Theorem 2 *The optimality conditions of the constrained variational problem (5) on $Y[t_s, t_f]^d$, i.e.*

$$J'(\mathbf{y}(t_f))\mathbf{w}(t_f) - \sum_{n=0}^{N-1} \int_{I_n} \dot{\mathbf{w}}(t) - \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}(t))\mathbf{w}(t) \, d\mathbf{A}(t) - \mathbf{l}^\top \mathbf{w}(t_s) = 0, \quad (20a)$$

$$- \sum_{n=0}^{N-1} \int_{I_n} \dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t)) \, d\mathbf{\Gamma}(t) = 0, \quad (20b)$$

$$-\mathbf{r}^\top [\mathbf{y}(t_s) - \mathbf{y}_s] = 0, \quad (20c)$$

$$\forall(\mathbf{w}, \mathbf{\Gamma}, \mathbf{r}) \in Y[t_s, t_f]^d \times \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d,$$

possess a unique solution $(\mathbf{y}, \mathbf{A}, \mathbf{l})$ in $Y[t_s, t_f]^d \times \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d$ that coincides with the solution of (15).

We start with considering the nominal equations (20c)-(20b).

Lemma 3 *The solution $\mathbf{y}(t)$ of (15c)-(15b) solves the extended variational formulation (20c)-(20b).*

Proof Let $\mathbf{y}(t)$ be the solution of (15c)-(15b). From $C^1[t_s, t_f]^d \subset Y[t_s, t_f]^d$ follows that $\mathbf{y} \in Y[t_s, t_f]^d$. Since the integral $\int_{t_s}^{t_f} g_i(t) d\Gamma_i(t)$ with $g_i(t) := \dot{\mathbf{y}}_i(t) - \mathbf{f}_i(t, \mathbf{y}(t))$ exists, also the integrals over the subintervals $\int_{t_n}^{t_{n+1}} g_i(t) d\Gamma_i(t)$ exist and it holds [21, Ch.VIII§6]

$$\int_{t_s}^{t_f} g_i(t) d\Gamma_i(t) = \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} g_i(t) d\Gamma_i(t) = \sum_{n=0}^{N-1} \int_{I_n} g_i(t) d\Gamma_i(t)$$

where the second equality is due to the extension (19) of the Riemann-Stieltjes integral, $i = 1, \dots, d$. Thus, equation (15b) becomes $\forall \mathbf{\Gamma} \in \text{NBV}[t_s, t_f]^d$

$$0 = \int_{t_s}^{t_f} \dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t)) d\mathbf{\Gamma}(t) = \sum_{n=0}^{N-1} \int_{I_n} \dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t)) d\mathbf{\Gamma}(t)$$

which coincides with (20b). \square

Lemma 4 *The extended variational formulation (20b)-(20c) possesses a unique solution $\mathbf{y}(t)$.*

Proof Let $\mathbf{y}(t)$ be a solution of (20b)-(20c). The space $\text{NBV}[t_s, t_f]^d$ contains, in particular, the functions that vanish everywhere except on (t_n, t_{n+1}) . Thus, a necessary condition for $\mathbf{y}(t)$ being a solution of (20b)-(20c) is that each addend has to vanish, i.e. $\int_{I_n} \dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t)) d\mathbf{\Gamma}(t) = 0 \quad \forall \mathbf{\Gamma} \in \text{NBV}(I_n)^d$ with $\mathbf{\Gamma}(t_{n+1}) = \mathbf{0}$. The fundamental theorem of variational calculus yields $\dot{\mathbf{y}}(t) - \mathbf{f}(t, \mathbf{y}(t)) = \mathbf{0}$ on (t_n, t_{n+1}) for all $n = 0, \dots, N-1$. On the other hand, $\text{NBV}[t_s, t_f]^d$ contains also the constant functions having a single jump in t_n . They give the necessary conditions $\dot{\mathbf{y}}(t_n) - \mathbf{f}(t_n, \mathbf{y}(t_n)) = \mathbf{0}$ for $n = 1, \dots, N$. Since $\mathbf{f}(t, \mathbf{y})$ is continuous in both variables and $\mathbf{y} \in C^0[t_s, t_f]^d$, $\mathbf{y}(t)$ is necessarily continuously differentiable on $[t_s, t_f]$. Thus, every solution of (20b)-(20c) satisfies (15b)-(15c) which possesses a unique solution. \square

As conclusion of this lemma, the dependency of the solution of the extended variational formulation (20b)-(20c) on the input data is continuous and thus the problem is well-posed.

Now, we focus on the adjoint problem in extended variational formulation which is for a given $\mathbf{y}(t)$: Find $(\mathbf{A}, \mathbf{l}) \in \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d$ such that (20a) holds for all $\mathbf{w} \in Y[t_s, t_f]^d$.

Lemma 5 *For the solution $\mathbf{y}(t)$ of (20b)-(20c), the corresponding adjoint solution $(\mathbf{A}, \mathbf{l}) \in \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d$ of (20a) is provided by (16).*

Proof We proceed in the same way as in the proof of Lemma 1, but choose $\mathbf{w} \in Y[t_s, t_f]^d$ for the multiplication and split the integral in (17) using the subintervals I_n (same arguments as in the proof of Lemma 3). Integration by parts of all integrals yields the equivalent equation

$$-\lambda^\top(t_s)\mathbf{w}(t_s) - \sum_{n=0}^{N-1} \int_{I_n} \lambda^\top(t) [\dot{\mathbf{w}}(t) - \mathbf{f}_y(t, \mathbf{y}(t))\mathbf{w}(t)] dt + J'(\mathbf{y}(t_f))\mathbf{w}(t_f) = 0.$$

Thus, the choice (16) provides a solution of (20a). \square

Lemma 6 *For the solution $\mathbf{y}(t)$ of (20b)-(20c), the corresponding adjoint solution $(\mathbf{A}, \mathbf{l}) \in \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d$ of (20a) is unique.*

Proof We follow mainly the proof of Lemma 2. Equation (20a) is equivalent to

$$\underbrace{\sum_{n=0}^{N-1} \int_{I_n} \dot{\mathbf{w}}(t) - \mathbf{f}_y(t, \mathbf{y}(t))\mathbf{w}(t) d\mathbf{A}(t) + \mathbf{l}^\top \mathbf{w}(t_s)}_{=: \hat{\mathbf{A}}(\mathbf{A}, \mathbf{l})(\mathbf{w})} = \underbrace{J'(\mathbf{y}(t_f))\mathbf{w}(t_f)}_{=: B(\mathbf{w})} \quad \forall \mathbf{w} \in Y[t_s, t_f]^d$$

where $\hat{\mathbf{A}}(\mathbf{A}, \mathbf{l})$ is also a linear functional on $Y[t_s, t_f]^d$ and $\hat{\mathbf{A}} : \text{NBV}[t_s, t_f]^d \times \mathbb{R}^d \rightarrow (Y[t_s, t_f]^d)'$ is linear in (\mathbf{A}, \mathbf{l}) . We show again that $\mathcal{N}(\hat{\mathbf{A}}) = \{(\mathbf{0}, \mathbf{0})\}$. Since $C^1[t_s, t_f]^d \subset Y[t_s, t_f]^d$, \mathbf{l} has to vanish due to the same arguments as used in the proof of Lemma 2. Thus, the following equation

$$\hat{\mathbf{A}}(\mathbf{A}, \mathbf{0})(\mathbf{w}) = \sum_{n=0}^{N-1} \int_{I_n} \dot{\mathbf{w}}(t) - \mathbf{f}_y(t, \mathbf{y}(t))\mathbf{w}(t) d\mathbf{A}(t) = 0 \quad \forall \mathbf{w} \in Y[t_s, t_f]^d$$

has to be satisfied also for $\mathbf{w} \in C^1[t_s, t_f]^d \subset Y[t_s, t_f]^d$, i.e. with $\mathbf{g}(t) := \dot{\mathbf{w}}(t) - \mathbf{f}_y(t, \mathbf{y}(t))\mathbf{w}(t)$ it becomes

$$\sum_{n=0}^{N-1} \int_{I_n} \mathbf{g}(t) d\mathbf{A}(t) = 0 \quad \forall \mathbf{g} \in C^0[t_s, t_f]^d.$$

Furthermore, as $\mathbf{g}(t)$ is continuous the integral $\int_{t_s}^{t_f} \mathbf{g}(t) d\mathbf{A}(t)$ exists and coincides with the sum of the integrals over the subintervals (same arguments as in the proof of Lemma 3) and the proof can be finished in the same way as that of Lemma 2. \square

With all this at hand we are able to prove Theorem 2.

Proof (of Theorem 2) Lemma 3 and 4 prove the existence of a unique solution of (20b)-(20c) coinciding with the solution of (15b)-(15c). For this solution, equation (20a) has a unique solution given by (16) due to Lemma 5 and 6. \square

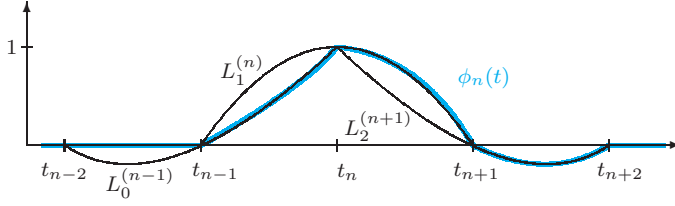


Fig. 2 Basis function ϕ_n of $Y_{\mathcal{P}}[t_s, t_f]^d$ with $k_0 = 1$, $k_n = 2$ for $n > 0$ and constant stepsizes $h_n = h$ for all n .

5 Petrov-Galerkin discretization of the extended optimality conditions

In order to solve the infinite-dimensional optimality conditions (20) numerically, the infinite-dimensional function spaces have to be approximated by finite-dimensional subspaces, the finite element spaces. This so-called *Petrov-Galerkin approximation* transfers the infinite-dimensional conditions into a finite-dimensional system of equations which can be solved on a computer. The first part of the section focuses on the finite-dimensional subspace, and the second part is devoted to the resulting system of equations.

5.1 Finite element spaces

This section deals with the discretization of the function spaces $Y[t_s, t_f]^d$ and $\text{NBV}[t_s, t_f]^d$ by choosing appropriate sets of basis functions.

Trial space To discretize the trial space $Y[t_s, t_f]^d$ we use piecewise polynomials of order k_n on the subinterval I_n

$$Y_{\mathcal{P}}[t_s, t_f]^d := \left\{ \mathbf{y} \in C^0[t_s, t_f]^d : \mathbf{y}|_{I_n} \in \mathcal{P}^{(k_n)}(I_n)^d \right\}. \quad (21)$$

We choose local basis functions ϕ_n that are composed of the fundamental Lagrangian polynomials (8) restricted to the particular subinterval. Figure 2 shows the basis function $\phi_n \in Y_{\mathcal{P}}[t_s, t_f]^d$ with $k_0 = 1$, $k_n = 2$ for $n > 0$ and $h_n = h$ for all n . The support of a single basis function depends on the orders and contains at most seven adjacent subintervals as BDF methods are stable up to order 6.

The solution $\mathbf{y} \in Y[t_s, t_f]^d$ is then approximated by

$$\mathbf{y}(t) \approx \mathbf{y}^h(t) := \mathbf{y}_s \phi_0(t) + \sum_{n=1}^N \mathbf{y}_n \phi_n(t)$$

which results in $N \cdot d$ degrees of freedom $\{\mathbf{y}_n \in \mathbb{R}^d\}_{n=1}^N$, since the initial value $\mathbf{y}_0 = \mathbf{y}_s$ is already fixed. To achieve locally the order $k_n > 1$, former values $\mathbf{y}_{n+1-k_n}, \dots, \mathbf{y}_n$ are reused to set up the interpolation polynomial of order k_n which is afterwards restricted to I_n .

Test space We approximate the test space $\text{NBV}[t_s, t_f]^d$ using Heaviside functions as basis functions. We choose them to be continuous from the right with discontinuity in t_n . Thus, a function $\mathbf{A} \in \text{NBV}[t_s, t_f]^d$ is approximated by the linear combination of these basis functions in the form

$$\mathbf{A}(t) \approx \mathbf{A}^h(t) := \sum_{n=1}^N h_{n-1} \boldsymbol{\lambda}_n H_n(t) \quad (22)$$

where the h_{n-1} appear for reasons which will become clear later. Note that \mathbf{A}^h is a step function with initial value $\mathbf{A}^h(t_s) = \mathbf{0}$ and jumps of magnitude $h_{n-1} \boldsymbol{\lambda}_n$ at t_n for $n = 1, \dots, N$. Thus, it is $\mathbf{A}^h(t_n) = \mathbf{A}^h(t_{n-1}) + h_{n-1} \boldsymbol{\lambda}_n$ at the time points and $\mathbf{A}^h(t) = \mathbf{A}^h(t_n)$ for inner points $t \in (t_n, t_{n+1})$. We denote this space by $Z_H[t_s, t_f]^d$.

Regarding the relation (16) between the adjoint solutions $\boldsymbol{\lambda}$ and \mathbf{A} , the classical derivative of \mathbf{A}^h fails to exist. But \mathbf{A}^h is still differentiable in a weak form such that its weak derivative is given by the Dirac measures at $\{t_1, \dots, t_N\}$ with heights $\{h_0 \boldsymbol{\lambda}_1, \dots, h_{N-1} \boldsymbol{\lambda}_N\}$, see e.g. [5, Sec. 4.24].

5.2 Finite-dimensional optimality conditions

In this section, we approximate the infinite-dimensional optimality conditions (20) by finite-dimensional equations that result from approximating the function spaces by the finite element spaces of Section 5.1. The resulting system of equations will be discussed in the following.

Theorem 3 *The discretized optimality conditions, i.e.*

$$J'(\mathbf{y}^h(t_f)) \mathbf{w}^h(t_f) - \sum_{n=0}^{N-1} \int_{I_n} \dot{\mathbf{w}}^h(t) - \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}^h(t)) \mathbf{w}^h(t) d\mathbf{A}^h(t) - [\mathbf{l}^h]^\top \mathbf{w}^h(t_s) = 0, \quad (23a)$$

$$- \sum_{n=0}^{N-1} \int_{I_n} \dot{\mathbf{y}}^h(t) - \mathbf{f}(t, \mathbf{y}^h(t)) d\mathbf{\Gamma}^h(t) = 0, \quad (23b)$$

$$-[\mathbf{r}^h]^\top [\mathbf{y}^h(t_s) - \mathbf{y}_s] = 0, \quad (23c)$$

$$\forall (\mathbf{w}^h, \mathbf{\Gamma}^h, \mathbf{r}^h) \in Y_{\mathcal{P}}[t_s, t_f]^d \times Z_H[t_s, t_f]^d \times \mathbb{R}^d,$$

are equivalent to the BDF scheme (7) with prescribed stepsizes and orders together with its discrete adjoint scheme (10).

The above theorem is the main result of this section. The proof follows directly from the two lemmas given below.

Lemma 7 *The equations (23b)-(23c) are equivalent to the BDF scheme (7) with prescribed stepsizes and orders.*

Proof We first consider one addend of (23b)

$$\begin{aligned} & \int_{I_n} \dot{\mathbf{y}}^h(t) - \mathbf{f}(t, \mathbf{y}^h(t)) d\mathbf{I}^h(t) \\ &= [\mathbf{I}^h(t_{n+1}) - \mathbf{I}^h(t_n)]^\top \{ \dot{\mathbf{y}}^h(t_{n+1}) - \mathbf{f}(t_{n+1}, \mathbf{y}^h(t_{n+1})) \} \\ &= \gamma_{n+1}^\top \left\{ \sum_{i=0}^{k_n} \underbrace{h_n \dot{\phi}_{n+1-i}(t_{n+1})}_{=\alpha_i^{(n)}} \mathbf{y}_{n+1-i} - h_n \mathbf{f}(t_{n+1}, \mathbf{y}_{n+1}) \right\} \end{aligned}$$

where the first equality holds due to the extended Riemann-Stieltjes integral (19) in vector-valued version with coefficients $h_n \gamma_{n+1}$ of \mathbf{I}^h in (22). The second equality uses the properties of the basis functions ϕ_n . Here the appearance of the h_n in the coefficients of \mathbf{A}^h given by (22) becomes clear. Thus, (23b) can be written as a system of equations that is nonlinear in $\{\mathbf{y}_n\}_{n=1}^N$ and linear in $\gamma^\top := [\gamma_1^\top \ \gamma_2^\top \ \cdots \ \gamma_N^\top] \in (\mathbb{R}^d)^N$

$$\gamma^\top \left[(\mathbf{A} \otimes \mathbf{I}) \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{pmatrix} + \begin{pmatrix} \alpha_1^{(0)} \mathbf{y}_s \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \begin{pmatrix} h_0 \mathbf{f}(t_1, \mathbf{y}_1) \\ h_1 \mathbf{f}(t_2, \mathbf{y}_2) \\ \vdots \\ h_{N-1} \mathbf{f}(t_N, \mathbf{y}_N) \end{pmatrix} \right] = 0, \quad \forall \gamma \quad (24)$$

where $\mathbf{A} \otimes \mathbf{I}$ denotes the Kronecker tensor product, i.e. the $(N \cdot d) \times (N \cdot d)$ matrix with $d \times d$ blocks $a_{ij} \mathbf{I}$, and the quadratic matrix \mathbf{A} is lower triangular with band structure

$$\mathbf{A} = \begin{pmatrix} \alpha_0^{(0)} & 0 & 0 & 0 & \cdots \\ \alpha_1^{(1)} & \alpha_0^{(1)} & 0 & 0 & \cdots \\ \vdots & & & & \\ \cdots & 0 & \alpha_{k_{N-1}}^{(N-1)} & \cdots & \alpha_0^{(N-1)} \end{pmatrix}.$$

Equation (24) holds if and only if the term in the squared brackets vanishes. Since \mathbf{A} is lower triangular, each \mathbf{y}_{n+1} is determined directly from $\mathbf{y}_s, \mathbf{y}_1, \dots, \mathbf{y}_n$ by the n th equation of the squared brackets term in (24) which coincides with the n th step of (7b). So, together with the equivalence between (7a) and (23c) the lemma is shown. \square

Lemma 8 *For the solution $\mathbf{y}^h(t)$ of (23b)-(23c), the equation (23a) is equivalent to the discrete adjoint scheme (10) of the nominal BDF scheme.*

Proof Analogously to the beginning of the proof of Lemma 7, each integral in (23a) is given by

$$\begin{aligned} & \int_{I_n} \dot{\mathbf{w}}^h(t) - \mathbf{f}_{\mathbf{y}}(t, \mathbf{y}^h(t)) \mathbf{w}^h(t) d\mathbf{A}^h(t) \\ &= \lambda_{n+1}^\top \left\{ \sum_{i=0}^{k_n} \alpha_i^{(n)} \mathbf{w}_{n+1-i} - h_n \mathbf{f}_{\mathbf{y}}(t_{n+1}, \mathbf{y}_{n+1}) \mathbf{w}_{n+1} \right\}. \end{aligned}$$

Thus, equation (23a) can be formulated equivalently in matrix form with $\mathbf{w}^\top := [\mathbf{w}_1^\top \mathbf{w}_2^\top \cdots \mathbf{w}_N^\top] \in (\mathbb{R}^d)^N$

$$\begin{aligned} & [\mathbf{0} \cdots \mathbf{0} \ J'(\mathbf{y}_N)] \mathbf{w} - (\alpha_1^{(0)} \boldsymbol{\lambda}_1 - \mathbf{l})^\top \mathbf{w}_0 \\ & - \boldsymbol{\lambda}^\top \left[\mathbf{A} \otimes \mathbf{I} - \begin{pmatrix} h_0 \mathbf{f}_y(t_1, \mathbf{y}_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & h_{N-1} \mathbf{f}_y(t_N, \mathbf{y}_N) \end{pmatrix} \right] \mathbf{w} = 0, \quad \forall \mathbf{w}_0, \mathbf{w} \end{aligned} \quad (25)$$

which is linear in both the variations \mathbf{w}_0, \mathbf{w} and the unknown $\boldsymbol{\lambda}$. The equivalent time-stepping scheme goes backwards in time starting with $J'(\mathbf{y}_N) - \alpha_0^{(N-1)} \boldsymbol{\lambda}_N^\top + h_{N-1} \boldsymbol{\lambda}_N^\top \mathbf{f}_y(t_N, \mathbf{y}_N) = 0$. Thus, (25) is equivalent to (10) which finishes the proof. \square

The necessary conditions for the well-posedness of (23b)-(23c) are stated in numerous textbooks on BDF methods, for example in [23, Ch.4§3]. With the Lipschitz constant L of $\mathbf{f}(t, \mathbf{y})$, the sequence of stepsizes and orders has to satisfy $|h_n/\alpha_0^{(n)} L| < 1$ in order to provide a unique solution $\mathbf{y}^h(t)$ of (23b)-(23c). The solution depends continuously on the input data due to the stability of the integration scheme. Since $\mathbf{f}_y(t, \mathbf{y})$ is bounded by L for all (t, \mathbf{y}) and h_n, k_n satisfy $|h_n/\alpha_0^{(n)} L| < 1$, the matrix in (25) is non-singular and thus (23a) possesses a unique weak adjoint solution $\boldsymbol{\Lambda}^h(t)$. The solution depends continuously on the input data $J'(\mathbf{y}_N)$ since the stability of the nominal integration scheme is carried over to the discrete adjoint scheme [22]. The well-posedness of (23a) can also be established using the derivation of the equivalent scheme (10) by automatic differentiation of (7), cf. Section 3.

6 Convergence analysis of classical adjoints and weak adjoints

In this section, we focus on the asymptotic behavior of the solutions of the discrete adjoint scheme (10). Therefore, we consider a nominal BDF method of constant order k with constant stepsizes h using a self-starting procedure for $\mathbf{y}_1, \dots, \mathbf{y}_m$ with $m \geq k - 1$ fixed. We will call this a *non-adaptive BDF method*. As seen in Section 3.2, the main part of the discrete adjoint scheme, i.e. equation (10b) with $n = N - k, \dots, m$, is a consistent method of order k for a variant of the adjoint equation (2). However, the adjoint initialization and termination steps do not give consistent approximations. Nevertheless, we will prove that the approximations in the main part converge linearly to the exact classical solution $\boldsymbol{\lambda}(t)$ of (2) around the exact nominal solution $\mathbf{y}(t)$. Using this result, we then show the strong convergence of the finite element approximation $\boldsymbol{\Lambda}^h(t)$ towards the solution $\boldsymbol{\Lambda}(t)$ of (15a), i.e. to the weak solution of (2), in the total variation norm of $\text{NBV}[t_s, t_f]^d$.

6.1 Convergence of the discrete adjoints to the classical adjoint

The discrete adjoint scheme (10) of a non-adaptive BDF scheme reads

$$\alpha_0 \boldsymbol{\lambda}_N - J'(\mathbf{y}_N)^\top = h \mathbf{f}_y^\top(t_N, \mathbf{y}_N) \boldsymbol{\lambda}_N \quad (26a)$$

$$\sum_{i=0}^{N-1-n} \alpha_i \boldsymbol{\lambda}_{n+1+i} = h \mathbf{f}_y^\top(t_{n+1}, \mathbf{y}_{n+1}) \boldsymbol{\lambda}_{n+1}, \quad n = N-2, \dots, N-k \quad (26b)$$

$$\sum_{i=0}^k \alpha_i \boldsymbol{\lambda}_{n+1+i} = h \mathbf{f}_y^\top(t_{n+1}, \mathbf{y}_{n+1}) \boldsymbol{\lambda}_{n+1}, \quad n = N-k-1, \dots, m \quad (26c)$$

$$\sum_{i=0}^k \alpha_i^{(n+i)} \boldsymbol{\lambda}_{n+1+i} = h \mathbf{f}_y^\top(t_{n+1}, \mathbf{y}_{n+1}) \boldsymbol{\lambda}_{n+1}, \quad n = m-1, \dots, 0 \quad (26d)$$

where (26d) accounts for the nominal starting procedure. To investigate the scheme (26) purely as an integration method for the adjoint differential equation (2), we consider a continuously differentiable approximation $\tilde{\mathbf{y}}(t)$ satisfying $\tilde{\mathbf{y}}(t_n) = \mathbf{y}_n$ for $n = 0, \dots, N$, for example a quadratic spline function interpolating $\{\mathbf{y}_n\}_{n=0}^N$ and $\{\mathbf{f}(t_n, \mathbf{y}_n)\}_{n=0}^N$. With the adjoint differential equation around $\tilde{\mathbf{y}}(t)$

$$\dot{\tilde{\boldsymbol{\lambda}}}(t) = -\mathbf{f}_y^\top(t, \tilde{\mathbf{y}}(t)) \tilde{\boldsymbol{\lambda}}(t), \quad \tilde{\boldsymbol{\lambda}}(t_f) = J'(\tilde{\mathbf{y}}(t_f))^\top \quad (27)$$

the main steps (26c) can be seen as a BDF method of order k applied to (27). The adjoint initialization steps (26a)-(26b) can be interpreted as a starting procedure for (26c) giving inconsistent start values $\boldsymbol{\lambda}_N, \dots, \boldsymbol{\lambda}_{N-k+1}$.

In the following, we study the asymptotic behavior for decreasing $h \rightarrow 0$ and a fixed time point t^* which belongs to refining grids, i.e. for every stepsize h there exists an $n = n(h)$ such that $t^* = t_n$. The interval $[t_{m+1}, t_{N-k}]$ of the main part of (26) increases and approaches (t_s, t_f) for $h \rightarrow 0$. By $\|\cdot\|$ we denote any vector norm in \mathbb{R}^d .

Lemma 9 *Let $\mathbf{f}_y(t, \tilde{\mathbf{y}}(t))$ be continuously differentiable in $t \in [t_s, t_f]$ and $\tilde{\mathbf{y}}(t_n) = \mathbf{y}_n$ for $n = 0, \dots, N$ where $\{\mathbf{y}_n\}_{n=0}^N$ is computed by the non-adaptive BDF method of order k with constant stepsize h . Let $\tilde{\boldsymbol{\lambda}}(t)$ be the exact solution of the adjoint differential equation (27) and let $\{\boldsymbol{\lambda}_n\}_{n=1}^N$ be computed by (26). Then, for a fixed timepoint $t_n = t \in (t_s, t_f)$ there exists $H > 0$ such that*

$$\left\| \boldsymbol{\lambda}_n - \tilde{\boldsymbol{\lambda}}(t_n) \right\| = \mathcal{O}(h)$$

as the grid is refined with $H > h \rightarrow 0$.

Proof To ease the notion, we consider a scalar initial value problem, i.e. $d = 1$. Nevertheless, the proof is also valid for systems of initial value problems. Furthermore, we define some abbreviations $B(t) := \mathbf{f}_y^\top(t, \tilde{\mathbf{y}}(t))$ and $\eta := J'(\tilde{\mathbf{y}}(t_f))^\top$.

Thus, the starting procedure (26a)-(26b) can be written equivalently using $\lambda^\top := [\lambda_N \cdots \lambda_{N-k+1}]$ and the $k \times 1$ unit vector \mathbf{e}_1

$$\left[\tilde{\mathbf{A}} - h\mathbf{B}(t_N, h) \right] \lambda = \mathbf{e}_1 \eta$$

where $\tilde{\mathbf{A}} = \bar{\mathbf{I}} [\mathbf{A}_{N-k+1:N, N-k+1:N}]^\top \bar{\mathbf{I}}$ for the reverse identity matrix $\bar{\mathbf{I}}$ and the matrix \mathbf{A} from page 17, and

$$\mathbf{B}(t_N, h) := \begin{pmatrix} B(t_N) & & 0 \\ & \ddots & \\ 0 & & B(t_N - (k-1)h) \end{pmatrix} = B(t_N)\mathbf{I} + \mathcal{O}(h) \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & 1 \end{pmatrix}$$

using the Taylor series expansion of the entries $B(t_N - ih)$ around t_N . The matrix $\tilde{\mathbf{A}}$ is nonsingular since $\alpha_0 \neq 0$. Furthermore, for h small enough to satisfy $\|h\tilde{\mathbf{A}}^{-1}\mathbf{B}(t_N, h)\| < 1$ we can use the Neumann series to express the inverse of $\mathbf{I} - h\tilde{\mathbf{A}}^{-1}\mathbf{B}(t_N, h)$, see for example [25, Sec. II.1], which yields

$$\begin{aligned} \lambda &= \left[\tilde{\mathbf{A}} \left(\mathbf{I} - h\tilde{\mathbf{A}}^{-1}\mathbf{B}(t_N, h) \right) \right]^{-1} \mathbf{e}_1 \eta = \left\{ \sum_{j=0}^{\infty} \left(h\tilde{\mathbf{A}}^{-1}\mathbf{B}(t_N, h) \right)^j \right\} \tilde{\mathbf{A}}^{-1} \mathbf{e}_1 \eta \\ &= \left\{ \mathbf{I} + h\tilde{\mathbf{A}}^{-1}\mathbf{B}(t_N, h) + \mathcal{O}(h^2) \right\} \tilde{\mathbf{A}}^{-1} \mathbf{e}_1 \eta \\ &= \tilde{\mathbf{A}}^{-1} \mathbf{e}_1 \eta + h\tilde{\mathbf{A}}^{-1}\mathbf{B}(t_N)\tilde{\mathbf{A}}^{-1} \mathbf{e}_1 \eta + \mathcal{O}(h^2). \end{aligned} \quad (28)$$

We want to apply Theorem 4.3 of [15] to the linear differential equation (27). Note that the starting procedure satisfies the assumptions of the theorem due to (28). As BDF methods are strongly stable, the only essential root of the characteristic polynomial $\rho(z) = \sum_{i=0}^k \alpha_i z^{k-i}$ is the principal root $z_1 = 1$. Thus, Theorem 4.3 of [15] gives for certain constants K_1 and K_2

$$\lambda_n - \tilde{\lambda}(t_n) = \exp \left(\int_{t_f}^{t_n} -B(\tau) d\tau \right) \delta_1 + \theta \left(K_1 + \frac{K_2}{t_n - h - t_f} \right) h$$

where $|\theta| < 1$ in the scalar case ($\|\theta\| < 1$ for $d > 1$). The quantity δ_1 is

$$\delta_1 := \frac{1}{\rho'(1)} \sum_{i=0}^{k-1} \gamma_i (\lambda_{N-i} - \eta), \text{ where } \sum_{i=0}^{k-1} \gamma_i z^i := \frac{\rho(z)}{z-1}$$

and the coefficients γ_i sum up to 1, i.e. $\sum_{i=0}^{k-1} \gamma_i = 1$. The latter fact together with equation (28) gives for $\gamma^\top := [\gamma_0 \cdots \gamma_{k-1}]$

$$\begin{aligned} \delta_1 &= \gamma^\top \lambda - \eta = \gamma^\top \left[\tilde{\mathbf{A}}^{-1} \mathbf{e}_1 \eta + h\tilde{\mathbf{A}}^{-1}\mathbf{B}(t_N)\tilde{\mathbf{A}}^{-1} \mathbf{e}_1 \eta + \mathcal{O}(h^2) \right] - \eta \\ &= \left[\gamma^\top \tilde{\mathbf{A}}^{-1} \mathbf{e}_1 - 1 \right] \eta + h\gamma^\top \tilde{\mathbf{A}}^{-1}\mathbf{B}(t_N)\tilde{\mathbf{A}}^{-1} \mathbf{e}_1 \eta + \mathcal{O}(h^2). \end{aligned}$$

The coefficient $\gamma^\top \tilde{\mathbf{A}}^{-1} \mathbf{e}_1 - 1$ of the first addend vanishes which can be verified easily for all BDF methods up to order 6. Thus, we obtain

$$\begin{aligned} \lambda_n - \tilde{\lambda}(t_n) &= h \exp \left(\int_{t_n}^{t_f} B(\tau) d\tau \right) \gamma^\top \tilde{\mathbf{A}}^{-1} B(t_N) \tilde{\mathbf{A}}^{-1} \mathbf{e}_1 \eta \\ &\quad + h \theta \left(K_1 + \frac{K_2}{t_n - h - t_f} \right) + \mathcal{O}(h^2) \end{aligned}$$

where both coefficients are bounded which proves the assertion. \square

The main result of this subsection is the following.

Theorem 4 *Let $\mathbf{f}(t, \mathbf{y})$ be continuously differentiable with respect to (t, \mathbf{y}) . Let $\boldsymbol{\lambda}(t)$ be the exact solution of the adjoint differential equation (2) and let $\{\boldsymbol{\lambda}_n\}_{n=1}^N$ be computed by (26). Then, for a fixed timepoint $t_n = t \in (t_s, t_f)$ there exists $H > 0$ such that*

$$\|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}(t_n)\| = \mathcal{O}(h) \quad (29)$$

as the grid is refined with $H > h \rightarrow 0$.

Proof Let the continuously differentiable spline $\tilde{\mathbf{y}}(t)$ be composed of quadratic polynomials on I_n such that $\tilde{\mathbf{y}}(t_n) = \mathbf{y}_n$, $\tilde{\mathbf{y}}(t_{n+1}) = \mathbf{y}_{n+1}$ and $\dot{\tilde{\mathbf{y}}}(t_{n+1}) = \dot{\mathbf{f}}(t_{n+1}, \mathbf{y}_{n+1})$ for $n = 0, \dots, N-1$. Furthermore, we define the interpolation operator \mathcal{I} that maps a continuously differentiable function $\mathbf{g}(t)$ to a continuously differentiable spline $\mathcal{I}\mathbf{g}(t)$ that is composed of quadratic polynomials on I_n with $\mathcal{I}\mathbf{g}(t_n) = \mathbf{g}(t_n)$, $\mathcal{I}\mathbf{g}(t_{n+1}) = \mathbf{g}(t_{n+1})$ and $\dot{\mathcal{I}\mathbf{g}}(t_{n+1}) = \dot{\mathbf{g}}(t_{n+1})$ for $n = 0, \dots, N-1$. Then, the difference of $\tilde{\mathbf{y}}(t)$ and $\mathcal{I}\mathbf{y}(t)$ in C^0 -norm is

$$\|\tilde{\mathbf{y}}(t) - \mathcal{I}\mathbf{y}(t)\|_{C^0[t_s, t_f]^d} = \mathcal{O}(h)$$

using Taylor expansions and the convergence of the nominal BDF method. Due to the assumption on $\mathbf{f}(t, \mathbf{y})$, the exact nominal solution $\mathbf{y}(t)$ of (1) is twice continuously differentiable such that

$$\|\mathbf{y}(t) - \mathcal{I}\mathbf{y}(t)\|_{C^0[t_s, t_f]^d} = \mathcal{O}(h^2)$$

due to the approximation property of quadratic splines. Thus, it is

$$\|\tilde{\mathbf{y}}(t) - \mathbf{y}(t)\|_{C^0[t_s, t_f]^d} \leq \|\tilde{\mathbf{y}}(t) - \mathcal{I}\mathbf{y}(t)\|_{C^0} + \|\mathcal{I}\mathbf{y}(t) - \mathbf{y}(t)\|_{C^0} = \mathcal{O}(h). \quad (30)$$

Since both adjoint differential equations (2) and (27) are linear, their solutions $\boldsymbol{\lambda}(t)$ and $\tilde{\boldsymbol{\lambda}}(t)$ can be given explicitly. Subtracting the exact adjoint solutions and using (30) yields in the C^0 -norm

$$\left\| \tilde{\boldsymbol{\lambda}}(t) - \boldsymbol{\lambda}(t) \right\|_{C^0[t_s, t_f]^d} = \mathcal{O}(h) \quad (31)$$

which implies directly the pointwise convergence for every $t \in [t_s, t_f]$. Thus, together with Lemma 9 we obtain

$$\|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}(t_n)\| \leq \|\boldsymbol{\lambda}_n - \tilde{\boldsymbol{\lambda}}(t_n)\| + \|\tilde{\boldsymbol{\lambda}}(t_n) - \boldsymbol{\lambda}(t_n)\| = \mathcal{O}(h)$$

for $t_n \in (t_s, t_f)$. \square

Remark 1 If $\mathbf{f}(t, \mathbf{y})$ is k -times continuously differentiable in (t, \mathbf{y}) , the start errors of the nominal BDF method of order k are small enough (i.e. the convergence of order k is guaranteed), and the spline is of corresponding order, then (31) holds with order k in h .

The discrete adjoints resulting from the adjoint initialization and termination steps differ from the exact adjoints in a constant way. For $n = N, \dots, N - k + 1$ the difference is bounded by a positive constant c_n times the state $\boldsymbol{\lambda}(t_f) = J'(\mathbf{y}(t_f))^\top$, i.e.

$$\|\boldsymbol{\lambda}_n - \boldsymbol{\lambda}(t_n)\| \leq c_n \|J'(\mathbf{y}(t_f))\| + \mathcal{O}(h).$$

This can be shown using (31), the Taylor expansion of $\tilde{\boldsymbol{\lambda}}(t_n)$ around t_f and the Neumann series of the inverse of $\alpha_0^{(n)} \mathbf{I} - h \mathbf{f}_{\mathbf{y}}(t_{n+1}, \mathbf{y}_{n+1})$. For the discrete adjoints from the adjoint termination steps (26d), one also needs Lemma 9 and obtains a multiple of $\boldsymbol{\lambda}(t_s)$.

Without modifications of the adjoint initialization steps (26a)-(26b), the discrete adjoints on the main part converge linearly to the exact adjoint solution $\boldsymbol{\lambda}(t)$ of (2). Nevertheless, we still have to consider the oscillations of the discrete adjoints at the interval ends of $[t_s, t_f]$ which are due to the inconsistency of the adjoint initialization and termination steps. We will do this in the next section.

6.2 Convergence of the finite element approximation to the weak adjoint

We will prove the convergence of the finite element approximation of the weak adjoint to the exact weak adjoint of (2) given by (15a) with respect to the total variation norm of $\text{NBV}[t_s, t_f]^d$ (i.e. strong convergence).

Theorem 5 *The finite element approximation $\boldsymbol{\Lambda}^h(t) = \sum_{n=1}^N h_{n-1} \boldsymbol{\lambda}_n H_n(t)$ given by the discrete adjoint scheme (26) of a non-adaptive BDF method of constant order k with constant stepsize h converges to the exact weak adjoint solution $\boldsymbol{\Lambda}(t) = \int_{t_s}^t \boldsymbol{\lambda}(\tau) d\tau$ where $\boldsymbol{\lambda}(\tau)$ solves (2). The convergence is with respect to the total variation norm of $\text{NBV}[t_s, t_f]^d$.*

Proof Let $h := \frac{t_f - t_s}{N}$ be the stepsize of the equidistant grid. Thus, the nodes are $t_n = t_s + nh$ for $n = 0, \dots, N$. We use the norms mentioned in Section 4.1

and consider firstly the i th component, $1 \leq i \leq d$. To ease the notion, we set $\Lambda := \mathbf{\Lambda}_i$, $\Lambda^h := \mathbf{\Lambda}_i^h$, $g := \mathbf{g}_i$ such that the dual norm reads

$$\|\Lambda - \Lambda^h\|_{\text{NBV}[t_s, t_f]} = \sup_{\|g\|_{C^0[t_s, t_f]}=1} \left| \int_{t_s}^{t_f} g(t) d(\Lambda - \Lambda^h)(t) \right|.$$

As Λ is given by $\Lambda(t) = \int_{t_s}^t \lambda(\tau) d\tau$ and Λ^h is a jump function it holds [19, Sec.36 Example 3]

$$\int_{t_s}^{t_f} g(t) d(\Lambda - \Lambda^h)(t) = \int_{t_s}^{t_f} \lambda(t) g(t) dt - \sum_{n=1}^N h \lambda_n g(t_n).$$

Approximating the integral by the composite trapezoidal rule for equidistant grids yields

$$\begin{aligned} & h \left\{ \frac{1}{2} \lambda(t_0) g(t_0) + \sum_{n=1}^{N-1} \lambda(t_n) g(t_n) + \frac{1}{2} \lambda(t_N) g(t_N) \right\} + \mathcal{O}(h^2) - \sum_{n=1}^N h \lambda_n g(t_n) \\ &= h \left\{ \frac{1}{2} \lambda(t_0) g(t_0) + \sum_{n=1}^N [\lambda(t_n) - \lambda_n] g(t_n) - \frac{1}{2} \lambda(t_N) g(t_N) \right\} + \mathcal{O}(h^2). \end{aligned}$$

We obtain a bound for the $\text{NBV}[t_s, t_f]^d$ -dual norm of $\mathbf{\Lambda} - \mathbf{\Lambda}^h$ by taking the absolute value, using the triangle inequality and the fact that $\|g\|_{C^0[t_s, t_f]} = 1$, i.e.

$$\|\Lambda - \Lambda^h\|_{\text{NBV}[t_s, t_f]} \leq h \left\{ |\lambda(t_0)| + \sum_{n=1}^N |\lambda(t_n) - \lambda_n| + |\lambda(t_N)| \right\} + \mathcal{O}(h^2).$$

With Theorem 4 the sum over the main part becomes

$$\sum_{n=m+1}^{N-k} |\lambda(t_n) - \lambda_n| = \sum_{n=m+1}^{N-k} \mathcal{O}(h) = \mathcal{O}(1)$$

such that the norm is bounded by

$$\begin{aligned} & \|\Lambda - \Lambda^h\|_{\text{NBV}[t_s, t_f]} \\ & \leq h \left\{ |\lambda(t_0)| + \sum_{n=1}^m |\lambda(t_n) - \lambda_n| + \mathcal{O}(1) + \sum_{n=1}^{k-1} |\lambda(t_{N-n}) - \lambda_{N-n}| + |\lambda(t_N)| \right\} + \mathcal{O}(h^2). \end{aligned}$$

Since the magnitude of all remaining addends is bounded according to the end of Section 6.1 and their number is independent of the step number N , it is $\|\Lambda - \Lambda^h\|_{\text{NBV}[t_s, t_f]} = \mathcal{O}(h)$. As this holds for all $i = 1, \dots, d$ and the dual norm coincides with the total variation norm (cf. Section 4.1), the assertion is shown. \square

By small modifications in the proof of Theorem 5, the assertion can be widened to variable stepsizes in the starting procedure.

The uniform convergence in the total variation norm of $\text{NBV}[t_s, t_f]^d$ implies the pointwise convergence on the entire time interval which can be shown by utilizing the particular partition $\{t_s, \theta, t_f\}$ for an arbitrary time point $\theta \in [t_s, t_f]$. Thus, Theorem 5 implies the pointwise convergence of $\mathbf{A}^h(t)$ to $\mathbf{A}(t)$ on the entire time interval at least with the same convergence rate.

7 Numerical results

We illustrate the theoretical results with the help of a nonlinear test case with analytic nominal and adjoint solutions. The Catenary [12, p.15] is given by a second-order ODE

$$\ddot{y}(t) = p\sqrt{1 + \dot{y}(t)^2}, \quad p > 0.$$

We reformulate the initial value problem as system of first-order equations

$$\begin{aligned} \dot{y}_1(t) &= y_2(t) \\ \dot{y}_2(t) &= p\sqrt{1 + y_2(t)^2} \end{aligned}$$

and solve it on the interval $[0, 2]$ for $p = 3$ and $\mathbf{y}(0) = [1/3 \cosh(-3) \quad \sinh(-3)]^\top$. As criterion of interest we choose $J(\mathbf{y}(2)) = y_1(2)$. The analytic nominal solution is

$$\mathbf{y}(t) = \begin{pmatrix} B + \frac{1}{p} \cosh(pt + A) \\ \sinh(pt + A) \end{pmatrix}$$

and the analytic weak adjoint solution in the space $\text{NBV}[t_s, t_f]^2$ is

$$\mathbf{A}(t) = \begin{pmatrix} -\frac{1}{p^2} \ln(\cosh(pt + A)) + \frac{t}{p^2} \sinh(pt_f + A) \arctan(e^{pt+A}) \\ t \end{pmatrix} \quad (32)$$

where A and B are determined implicitly by the initial values.

7.1 Non-adaptive BDF method

We consider a non-adaptive BDF method of constant order 2 on an equidistant grid with stepsize h . The self-starting procedure consists of two first-order BDF steps with stepsize $h/2$. The simulations are performed in Matlab.

The lower row of Figure 3 compares the discrete adjoints for two different stepsizes $h = 2^{-4}$ and $h = 2^{-6}$ to the analytic solution of the adjoint differential equation. The oscillations of the discrete adjoints at the interval ends are due to the inconsistency of the adjoint initialization and termination steps of the discrete adjoint scheme with the adjoint differential equation (cf. Section 3.2). Nevertheless, the discrete adjoints converge on the open interval $(0, 2)$

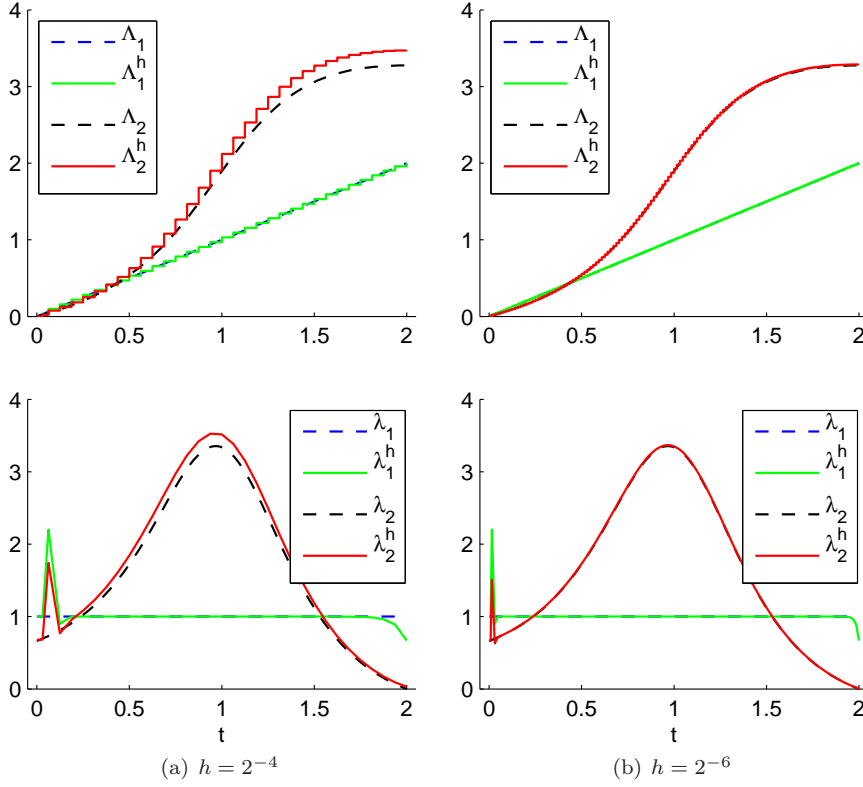


Fig. 3 Results of the non-adaptive BDF method for two different stepsizes. Comparison of the finite element approximation of the weak adjoint and the analytic weak adjoint (top) as well as the discrete adjoints in comparison to analytic Hilbert space adjoint (bottom) for different stepsizes.

towards the analytic adjoint solution as proven by Theorem 4. In the upper row of Figure 3 the finite element approximation $\mathbf{\Lambda}^h(t)$ is compared to the weak adjoint $\mathbf{\Lambda}(t)$ given by (32). It converges on the whole time interval as shown by Theorem 5.

Figure 4 shows the Euclidean norm of the difference between the analytic weak adjoint (32) and the finite element approximation, i.e.

$$\text{Error} = \|\mathbf{\Lambda}(t) - \mathbf{\Lambda}^h(t)\|_2,$$

evaluated at the final time $t = t_f = 2$ and at some interior time point $t = 1.25$, respectively, for shrinking stepsizes. The error evaluated at the final time decreases at second order rate, a somewhat better behavior than predicted by the convergence theory of Section 6.2. This might be due to the second order convergence of the discrete adjoints at the initial time together with a possible cancellation of discrepancies of the discrete adjoints at the interval ends (depicted in the lower row of Figure 3). Overall, this observation calls for

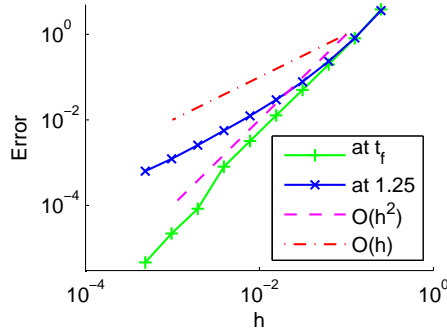


Fig. 4 Convergence of the finite element approximation of the weak adjoint to the analytic weak adjoint. Error evaluated at the final time $t_f = 2$ and at the interior time point $t = 1.25$.

a closer theoretical investigation. The error at the interior time point $t = 1.25$ shows the expected linear convergence, cf. Theorem 5 and the subsequent comment on the pointwise convergence.

7.2 Adaptive BDF method

The software package DAESOL-II [2] provides an efficient realization of a variable-order variable-stepsize BDF method based on a sophisticated order and stepsize selection. Furthermore, it contains efficient ways to compute the discrete adjoints [3, 4, 2]. We solved the Catenary for two different accuracies (relative tolerance 10^{-4} and 10^{-9}) to get a first asymptotic impression of the finite element approximation of the adjoint in the case of fully adaptive BDF methods. The results are depicted in Figure 5.

In areas of constant BDF order (fourth row of Figure 5) and constant step-sizes (third row), the discrete adjoints converge to the analytic adjoint solution (second row) as seen in the right column on the interval $(1, 1.7)$ approximately. On the other areas, i.e. where the order is varying and stepsize is changing, the discrete adjoints are highly oscillating (second row). Nevertheless, also in these cases, the finite element approximations $\mathbf{A}^h(t)$ converge to the analytic weak adjoint solution (32) on the entire time interval (first row of Figure 5).

8 Summary and outlook

In this contribution, we have addressed the issue of relating the discrete adjoints of variable-order variable-stepsize BDF methods to the solution of the adjoint differential equation (2). Since for multistep methods the common Hilbert space setting is not appropriate to interpret the discrete adjoints, we have developed a new Banach space approach. It is based on a constrained variational problem in the space of all continuously differentiable functions

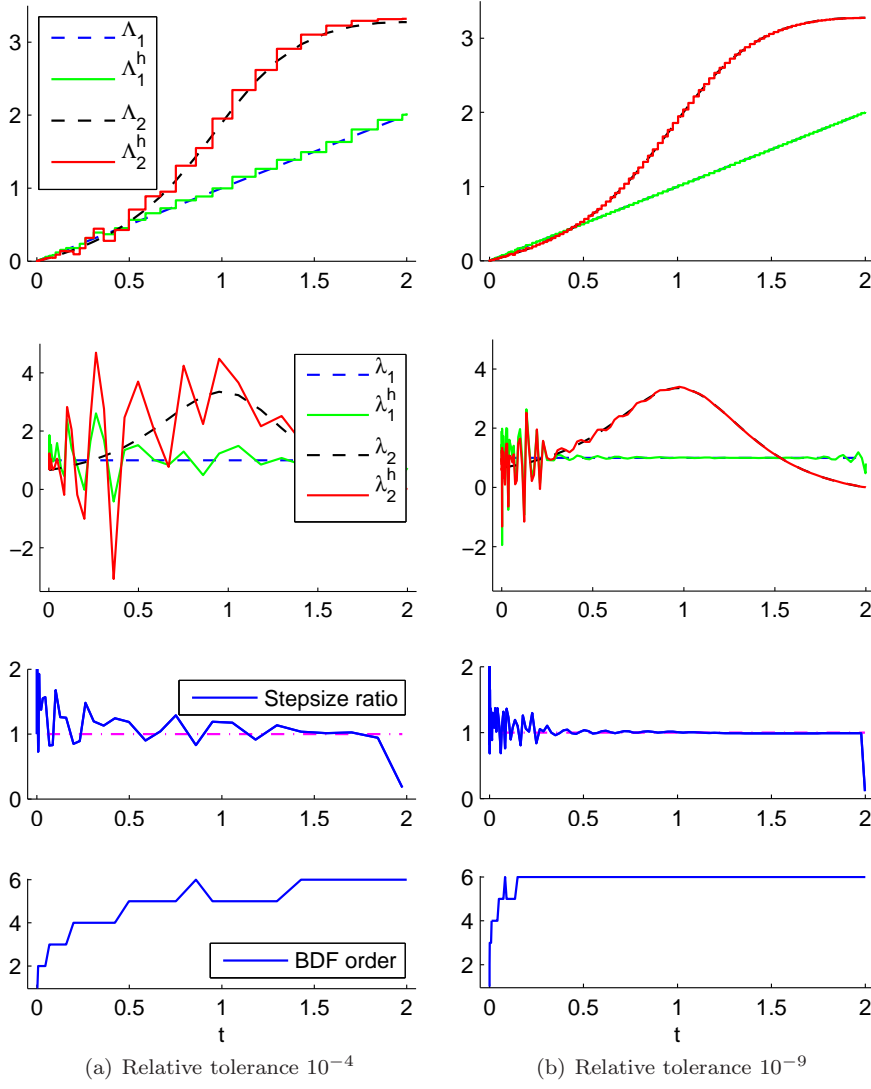


Fig. 5 Results of the adaptive BDF method for different accuracies. Comparison of the finite element approximation of the weak adjoint and the analytic weak adjoint (top) as well as the discrete adjoints in comparison to analytic Hilbert space adjoints (second row). Stepsize ratio (third row) and BDF order (bottom) of the integration scheme.

with Lagrange multiplier in the space of all normalized functions of bounded variation. We have approximated the infinite-dimensional optimality conditions by a Petrov-Galerkin discretization and have shown the equivalence of the resulting equations to the BDF scheme and its discrete adjoint scheme obtained by adjoint internal numerical differentiation. Thus, discretization and

optimization commute in the presented framework and the finite element approximation of the weak adjoint is obtained by a simple post-processing of the discrete adjoints. Furthermore, we have demonstrated that the discrete adjoint scheme of a non-adaptive BDF method produces discrete adjoints which converge linearly to the solution of (2) on the inner time interval although the adjoint initialization steps are inconsistent. We have used this result to prove the linear convergence of the finite element approximation on the entire time interval to the weak adjoint solution of (2) in the space of normalized functions of bounded variation.

The theoretical results have been observed numerically using a non-adaptive BDF method to solve the Catenary. Additionally, we have given numerical evidence that the finite element approximation serves as proper quantity to approximate the weak adjoint also in the case of fully adaptive BDF methods, i.e. also in areas of variable order and variable stepsize.

Thus, we now have a quantity at hand which can be used within global error estimation techniques. The functional-analytic framework allows to carry over estimation techniques from finite element methods to BDF methods. Furthermore, the approximations to the weak adjoints can now be computed efficiently and accurately by automatic differentiation of the efficient variable-order variable-stepsize BDF method without the need of explicit derivation of the adjoint equations.

Acknowledgements The authors express their gratitude to Christian Kirches and Andreas Potschka for valuable discussions on the subject. Scientific support of the DFG-Graduate-School 220 “Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences” is gratefully acknowledged. Funding graciously provided by the German Ministry of Education and Research (Grant ID: 03MS649A), and the Helmholtz association through the SBCancer programme. The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement n° FP7-ICT-2009-4 248940.

References

1. Adams, R., Fournier, J.: Sobolev Spaces, *Pure and Applied Mathematics (Amsterdam)*, vol. 140, second edn. Elsevier/Academic Press, Amsterdam (2003)
2. Albersmeyer, J.: Adjoint based algorithms and numerical methods for sensitivity generation and optimization of large scale dynamic systems. Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg (2010).
3. Albersmeyer, J., Bock, H. G.: Sensitivity Generation in an Adaptive BDF-Method. In: H. G. Bock, E. Kostina, X. Phu, R. Rannacher (eds.) *Modeling, Simulation and Optimization of Complex Processes: Proceedings of the International Conference on High Performance Scientific Computing*, March 6–10, 2006, Hanoi, Vietnam, pp. 15–24. Springer-Verlag Berlin Heidelberg (2008)
4. Albersmeyer, J., Bock, H.G.: Efficient sensitivity generation for large scale dynamic systems. Tech. rep., SPP 1253 Preprints, University of Erlangen (2009)
5. Alt, H. W.: Lineare Funktionalanalysis, 4 edn. Springer-Verlag Berlin Heidelberg (2002)
6. Berkovitz, L.: Optimal Control Theory, *Applied Mathematical Sciences*, vol. 12. Springer-Verlag, New York (1974)
7. Bock, H. G.: Numerical treatment of inverse problems in chemical reaction kinetics. In: K. Ebert, P. Deufhard, W. Jäger (eds.) *Modelling of Chemical Reaction Systems*,

-
- Springer Series in Chemical Physics*, vol. 18, pp. 102–125. Springer-Verlag, Heidelberg (1981).
8. Bock, H. G.: Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen, *Bonner Mathematische Schriften*, vol. 183. Universität Bonn, Bonn (1987).
 9. Bock, H. G., Plitt, K. J.: A Multiple Shooting algorithm for direct solution of optimal control problems. In: Proceedings of the 9th IFAC World Congress, pp. 242–247. Pergamon Press, Budapest (1984).
 10. Bock, H. G., Schlöder, J. P., Schulz, V.: Numerik großer Differentiell-Algebraischer Gleichungen – Simulation und Optimierung. In: H. Schuler (ed.) *Prozeßsimulation*, pp. 35–80. VCH Verlagsgesellschaft mbH, Weinheim (1994)
 11. Cao, Y., Li, S., Petzold, L.: Adjoint sensitivity analysis for differential-algebraic equations: algorithms and software. *Journal of Computational and Applied Mathematics* **149**, 171–191 (2002)
 12. Hairer, E., Nørsett, S., Wanner, G.: Solving Ordinary Differential Equations I, *Springer Series in Computational Mathematics*, vol. 8, second edn. Springer-Verlag, Berlin (1993)
 13. Hartman, P.: Ordinary differential equations, *Classics in Applied Mathematics*, vol. 38. SIAM, Philadelphia, PA (2002). Corrected reprint of the second (1982) edition [Birkhäuser, Boston, MA; MR0658490 (83e:34002)]
 14. Hartmann, R.: Adjoint consistency analysis of Discontinuous Galerkin discretizations. *SIAM J. Numer. Anal.* **45**, 2671–2696 (2007)
 15. Henrici, P.: Error Propagation for Difference Methods. Robert E. Krieger Publishing Co., Huntington, N. Y. (1970). Reprint of the 1963 edition
 16. Ioffe, A., Tihomirov, V.: Theory of extremal problems, *Studies in Mathematics and its Applications*, vol. 6. North-Holland Publishing Co., Amsterdam (1979).
 17. Johnson, C.: Numerical solutions of partial differential equations by the finite element method. Cambridge University Press, Cambridge (1987)
 18. Johnson, C.: Error estimates and adaptive time-step control for a class of one-step methods for stiff ordinary differential equations. *SIAM Journal on Numerical Analysis* **25**(4), 908–926 (1988).
 19. Kolmogorov, A., Fomin, S.: Introductory real analysis. Revised English edition. Translated from the Russian and edited by Richard A. Silverman. Prentice-Hall Inc., Englewood Cliffs, N.Y. (1970)
 20. Luenberger, D.: Optimization by vector space methods. Wiley Professional Paperback Series. John Wiley & Sons, Inc., New York, NY (1969).
 21. Natanson, I.: Theorie der Funktionen einer reellen Veränderlichen. Akademie-Verlag, Berlin (1975). Übersetzung nach der zweiten russischen Auflage von 1957, Herausgegeben von Karl Bögel, Vierte Auflage, Mathematische Lehrbücher und Monographien, I. Abteilung: Mathematische Lehrbücher, Band VI
 22. Sandu, A.: Reverse automatic differentiation of linear multistep methods. In: C. Bischof, H. Bücker, P. Hovland, U. Naumann, J. Utke (eds.) *Advances in Automatic Differentiation, Lecture Notes in Computational Science and Engineering*, vol. 64, pp. 1–12. Springer-Verlag, Berlin (2008)
 23. Shampine, L.: Numerical solution of ordinary differential equations. Chapman & Hall, New York (1994)
 24. Walther, A.: Automatic differentiation of explicit Runge-Kutta methods for optimal control. *Comput. Optim. Applic.* **36**, 83–108 (2007)
 25. Werner, D.: Funktionalanalysis. Springer-Verlag, Berlin (2000)
 26. Wloka, J.: Funktionalanalysis und Anwendungen. Walter de Gruyter, Berlin-New York (1971). De Gruyter Lehrbuch